FEBS Letters 589 (2015) 3005–3013

FEBS Letters

journal homepage: www.FEBSLetters.org

Review

On the demultiplexing of chromosome capture conformation data

Ivan Junier [a,b,c,*], Yannick G. Spill [b,d], Marc A. Marti-Renom [b,d], Miguel Beato [b,c], François le Dily [b,c]

[a] Laboratoire Adaptation et Pathogénie des Micro-organismes – UMR 5163, Université Grenoble 1, CNRS, BP 170, F-38042 Grenoble Cedex 9, France
[b] Gene Regulacion, Stem Cells, and Cancer Program, Centre de Regulaciò Genòmica (CRG), 08003 Barcelona, Spain
[c] Universitat Pompeu Fabra (UPF), Barcelona, Spain
[d] Genome Biology Group, Centre Nacional d'Anàlisi Genòmica (CNAG), 08028 Barcelona, Spain

ARTICLE INFO

Article history:
Received 27 April 2015
Revised 24 May 2015
Accepted 26 May 2015
Available online 6 June 2015

Edited by Wilhelm Just

Keywords:
Genomic contact maps
Demultiplexing
Deconvolution
Restrained-based methods
Polymer models of chromosomes

ABSTRACT

How to describe the multiple chromosome structures that underlie interactions among genome loci and how to quantify the occurrence of these structures in a cell population remain important challenges to solve, which can be addressed via a proper demultiplexing of chromosome capture conformation related data. Here, we first aim to review two main methodologies that have been proposed to tackle this problem: restrained-based methods, in which the resulting chromosome structures stem from the multiple solutions of a distance satisfaction problem; and thermodynamic-based methods, in which the structures stem from the simulation of polymer models. Next, we propose a novel demultiplexing method based on a matrix decomposition of contact maps. To this end, we extend the notion of topologically associated domains (TADs) by introducing that of statistical interaction domains (SIDs). SIDs can overlap and occur in a cell population at certain frequencies, and we propose a simple method to estimate these frequency values. As an application, we show that SIDs that measure 100 kb to tens of Mb long occur both frequently and specifically in the human genome.

1. Introduction

At scales beyond a few kilobase pairs (kb), most of our knowledge of chromosome structure has originated from the development of high-throughput molecular methods, which are also known as chromosome capture conformation (3C) techniques [1]; further reviewed in [2]. 3C-based techniques allow for the generation of matrices of interactions – also called contact maps (see Figures) – that gather contact frequencies between all pairs of chromosomal loci (or bins), where the loci are sorted according to their position along the chromosomes, with sizes that can be as small as 1 kb [3]. At the smallest scales, these contact maps, in combination with protein binding profiles, have revealed the ubiquitous presence of large protein complexes that can bind multiple loci, with CCCTC-binding factor (CTCF) and cohesin playing central roles [4,5]. Chromosomes in interphase are thus bridged in many places by loops that measure tens to hundreds kbp long and strongly correlate with the presence of CTCF binding sites [6,3]. However, all pairs of sites are not equivalent, which indicates that loops form preferentially between specific sites [3].

These loops lie at the core of the higher organization of chromosomes into the so-called topologically associated domains (TADs). TADs are contiguous genomic domains in which all pairs of loci interact frequently [7–10]. They can be easily identified from the contact maps based on the presence of contiguous blocks of interactions (see Figures). In mammals, they are typically 1 Mb long and are visible during interphase [11]. In addition to their high reproducibility [11], they are very similar between cell types, and many of them are conserved between distant species, e.g., from mouse to human [9,12]. Loops and TADs [13], together with CTCF and cohesin [14], are thus likely to play important roles in the proper functioning of cells, which supports the concept that the structure of chromosomes is a key aspect of cell functions [15,16].

Chromosomes are also known to be continuously shaped by transcription, by proteins that bind, wrap or stretch DNA, and by enzymes that cut or repair DNA. *Chromosomal loci are thus dynamical*, with Brownian-like types of motion that are constrained or facilitated by the local chromosome structure [17,18]. Due to the importance of the relative spatial positioning between *cis*-regulatory elements and genes in transcriptional regulation [19], these dynamical properties are believed to play an important role in the ability of chromosomal domains to respond differentially to

* Corresponding author at: Laboratoire Adaptation et Pathogénie des Micro-organismes – UMR 5163, Université Grenoble 1, CNRS, BP 170, F-38042 Grenoble Cedex 9, France.
E-mail address: ivan.junier@ujf-grenoble.fr (I. Junier).

http://dx.doi.org/10.1016/j.febslet.2015.05.049
0014-5793/

regulatory signals [24,25]. In particular, variations in the structural properties of chromosomes have been shown to be associated with variations in transcriptional properties [24]. This can be observed, for instance, by the reshaping of TADs during the sequential activation of *hox* genes in vertebrate limb development [27–29].

TADs and loops are also expected to be dynamical entities, as suggested in particular by the heterogeneity of single-cell contact maps obtained among a given cell population [30]. The major challenges of chromosome structuring thus include (i) identifying the multiple structures that are responsible for the contact maps and (ii) quantifying the occurrence of these structures. In the spirit of signal analysis, we refer to this problem as the *demultiplexing problem* (*demuxing problem* in short) – this has also been recently referred to as the "deconvolution problem" [24].

From an experimental point of view, precisely (e.g., at a 1-kb resolution) and exhaustively quantifying the dynamics and heterogeneities of chromosomal conformations raises extreme challenges. Demuxing methods have thus far primarily focused on the computational modeling of chromosomes. Along this line, "structural demuxing" approaches have consisted of generating chromosomal conformations such that the reconstructed contact maps coincide as much as possible with the input contact maps and, when possible, such that their statistical properties coincide as much as possible with other measurements of chromosome structuring (e.g., 3D-FISH).

The goal of this article is two-fold. First, we provide an overview of structural demuxing methods. Second, we propose a novel method that is rooted in the matrix analysis of contact maps, that is, without referring to any explicit chromosome conformation. Using this framework, we show that structural heterogeneities ranging from 100 kb to tens of Mb must be considered when describing the structure of human chromosomes. As a discussion, we present both technical and experimental challenges related to our novel demuxing method.

2. Structural demultiplexing approaches

In 3C-related methods, the number of contacts between any two bins in a population of cells is estimated by using a DNA ligation protocol that is combined with sequencing technology [1,2]. Accordingly, a contact between bins i and j (Fig. 1) is detectable if and only if i and j are sufficiently close in space to be cross-linked by formaldehyde. In this case, raw interaction data corresponds to the number of sequence reads that map both i and j. These read numbers are then normalized to mitigate biases that originate mostly from restriction enzyme cutting, GC content and sequence uniqueness [31] (see [32] for a recent review of normalization procedures). The resulting contact frequencies are proportional to the fraction of cells in which the bins are spatially close, with a coefficient of proportionality that is identical, in theory, for all pairs of bins.

As we review below, two classes of structural approaches have been developed to demux these contact maps. Both methods are aimed at generating an ensemble of spatial conformations of the chromosomes, whose statistical properties are compatible with input contact maps. *Restrained-based* (*RB*) *approaches* rely primarily on an implicit relationship between the contact frequencies and spatial distances between bins (Fig. 1). They are aimed at reconstructing the conformations by satisfying as many distance constraints as possible. Multiple solutions arise from the

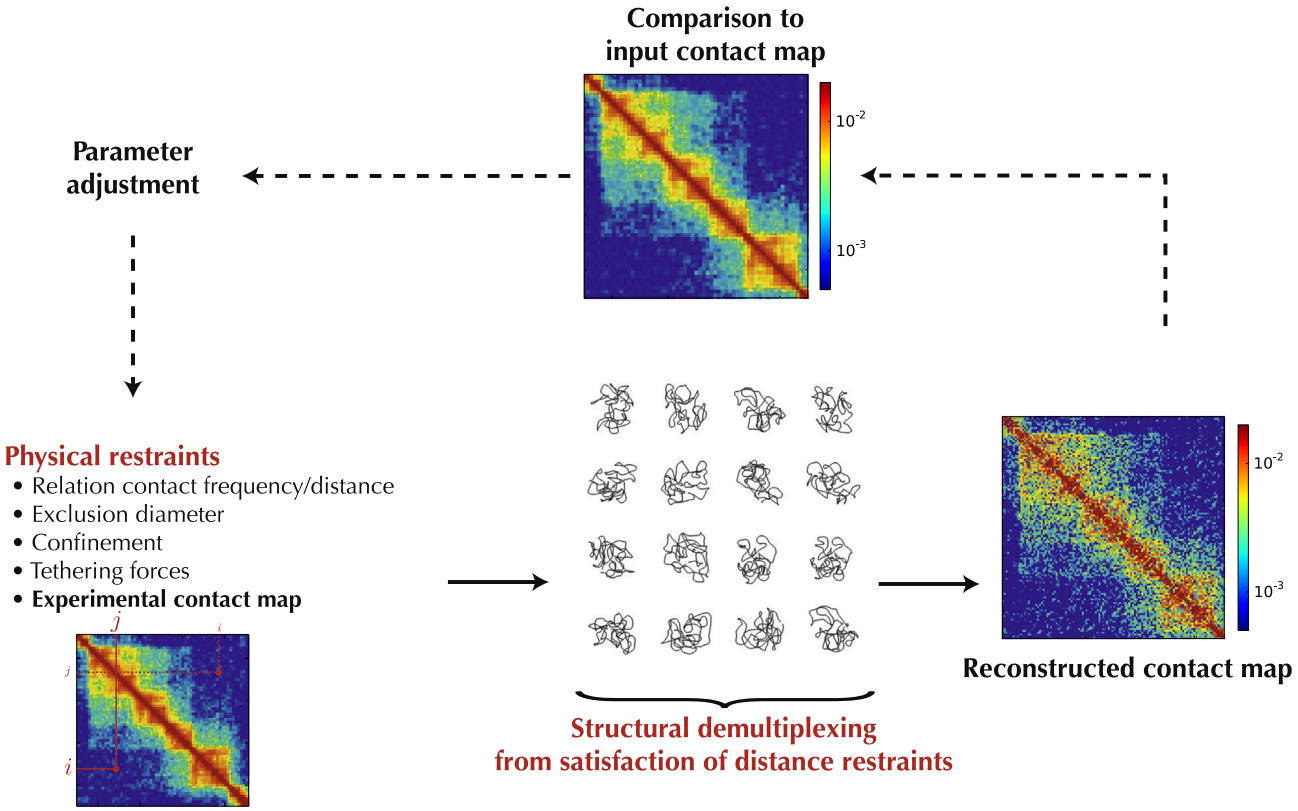

Fig. 1. Schematic representation of restraint-based (RB) framework. In RB modeling, experimental contact maps are explicitly used during an optimization procedure that involves satisfying distances that correspond to input contact frequencies. For that purpose, experimental frequencies are converted to target distances, frequently using an inverse power law (see text), and are used to constrain the modeling. Several conformers can then be extracted from the optimization procedure to provide a demultiplexing solution of the contact map as a population of conformations [20,21]. These conformers can then be used to reconstruct contact maps, which can further be compared with the experimental data to adjust the modeling parameters.

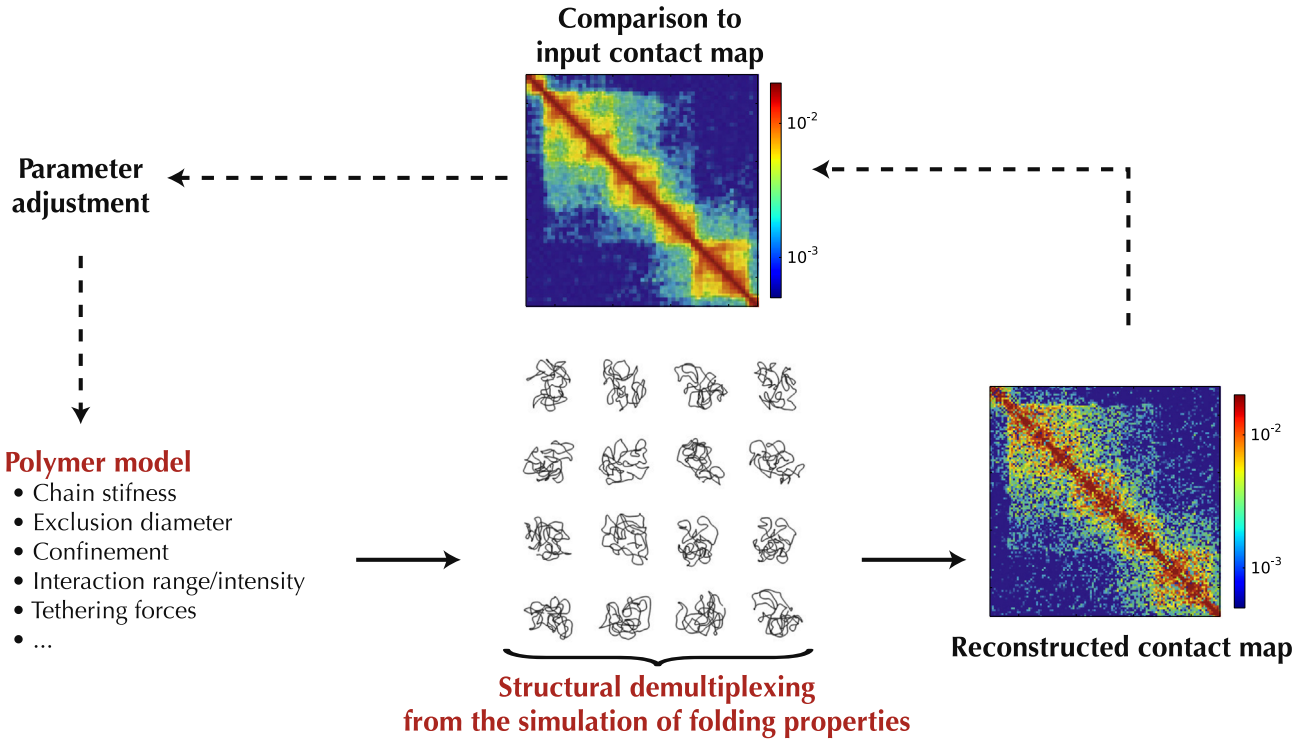

Fig. 2. Schematic representation of the thermodynamic-based (TB) framework. In TB modeling, a multitude of conformers are generated using a polymer model of the chromosomes with fixed parameters. Provided a distance below which two loci are considered to be in contact, contact pairs can be identified for every conformation, and the resulting statistics provides a reconstructed contact map. This contact map can then be compared with the input contact map, and parameters can be iteratively adjusted to eventually obtain a contact map that is as similar as possible to the input contact map [22–24].

impossibility of satisfying all pairwise distances using a single conformation and from the degeneracy of possible optimal solutions. In *thermodynamic-based* (*TB*) *approaches*, thermodynamic laws are used to simulate an ensemble of conformations given a predefined set of parameters of a polymer model (Fig. 2). The presence of different conformations is the natural consequence of entropy maximization (thermodynamic systems tend to explore as many conformations as possible) that is nevertheless constrained by the energy costs associated with each conformation, with a balance controlled by the physiological temperature. TB modeling has thus been described as "direct" modeling, whereas RB approaches have been described as "inverse" modeling [33] because conformations are deduced from experimental data "only" [34].

2.1. Restraint-based (RB) approaches

Data-driven RB approaches (Fig. 1) focus on translating data into structural features that may later be used to explain specific observations of the function of the modeled genomic domain or genome. For that purpose, chromatin is represented using a beads-on-a-string model. The position of the beads is modified to seek the minimum of a cost function, which can always be written as

$$E_{tot} = E_{poly} + E_{data}$$

E_{poly} is a cost function that penalizes deviations from the beads-on-a-string model. It favors the connectivity of neighboring beads and penalizes bead interpenetration. E_{data} is the dominant contribution in RB approaches and determines the agreement between the candidate conformation and the data, *e.g.*, HiC contact maps. A number of functional forms are used for E_{data}, and the most popular ones are harmonic or truncated harmonic potentials. Next, because HiC provides contact frequencies, it is necessary for RT approaches to convert them into distances to be able to generate conformations. Most methods use an empirical relationship, which was experimentally confirmed on a limited range of distances [35]:

$$f_{ij} \propto \frac{1}{d_{ij}^{a}} \quad a > 0$$

where f_{ij} is the observed contact frequency, d_{ij} is the target distance between loci i and j, and a is an exponent that is often set to 1. Optimization algorithms range from conjugate gradients to combinations of Brownian motion and simulated annealing.

The conformations resulting from RB approaches have been used to demux input interaction matrices into a set of differentiated conformational states. Two main types of frameworks have been proposed to date: (i) *a posteriori* frameworks, where 3D conformations are structurally compared to identify the most likely ones states, and (ii) *a priori* frameworks, where the search for solutions follows a certain probability distribution that is informed by the input matrices. Both methods are aimed at identifying differentiated conformational states that agree with 3C-based experiments, as we now explain in more detail.

2.1.1. A posteriori approaches

The first application of RB approaches in an attempt to demux HiC contact maps involved studying the structural variability of the resulting conformations of the α-globin genomic domain [36]. Hierarchical clustering of the conformations that resulted from the RB optimization procedure then revealed that the structural diversity was much lower in the inactive form of the α-globin domain than in the active form, which revealed that a larger complexity was inherited from the interaction matrix of the active form compared with the inactive form. Most importantly, the set of conformations revealed that the contact map obtained for the active form could only be explained by the existence of several different

states [36]; this is in contrast, for instance, to what is often expected in the problem of protein folding. Similar approaches based on a clustering of conformations that optimize a cost function have since then been investigated. For instance, Rousseau et al. used a Monte Carlo sampling strategy together with a hierarchical clustering method to model the conformational states of the HoxA gene cluster [37]. Similarly, Ay et al. reconstructed the *Plasmodium falciparum* genome by clustering 100 conformations that were obtained using a similar optimization procedure [38].

2.1.2. A priori approaches

In *a posteriori* approaches, 3D conformations are structurally compared to identify the most likely conformational states. In contrast, *a priori* frameworks follow a certain probability distribution that is informed by the input matrices. Using a Bayesian model for structural variability, the authors of "BACH-Mix" proposed two types of contact map demuxing [39]. Both types begin with a common step, called BACH, which is an RB-based approach for which the functional form of E_{data} is justified using Bayesian inference. Instead of repeating this step a number of times, as in the *a priori* approaches, the optimization is performed once to yield what is referred to as the consensus structure. The second step differs depending on whether one seeks to assess the continuous variability of a given structure or whether another structurally unrelated conformation is to be found. The former choice is called BACH-Mix because the assessment of variability relies on a Poisson mixture model. The latter consists of iteratively running BACH on residual HiC matrices, where contacts from the previous consensus structure are removed.

A more "population-based" demuxing RB approach relies on exploiting the fact that contact frequencies result from the superposition of a large number of different conformations [40,20]. Here, the computation consists of a unique simulation, which simultaneously generates a large number of plausible conformations. In this simulation, the cost function E_{tot} is a function of a large number of conformers, which are all optimized at the same time. During the simulation, the cost function enforces contacts between two loci in only a subset of all conformers. Using this approach, contact maps obtained in yeast [20] and in humans [40] have been demuxed by creating an ensemble of structures whose diversity mimicked that of the cells used to perform the experiments.

2.2. Thermodynamics-based (TB) approaches

2.2.1. Using polymer models for generating chromosomal conformations

In TB approaches, in vivo chromosomes are modeled as single polymer chains that include at least three fundamental features: (i) the chains can bend more or less easily, depending on their persistence length, (ii) the chains cannot overlap due to the physical exclusion properties of chromatin, and (iii) the chains are confined to a volume such that the resulting DNA density is similar to that of nuclei in vivo. If necessary, properties (i) and (ii) can be made more complex by further considering heterogeneous properties along the chains. Interactions between specific sites or between random sites can also be considered to mimic the presence of bridging complexes or the effect of condensation forces. In all cases, the parameters are fixed, and the contact maps are calculated by simulating the folding properties of the chains according to thermodynamic laws. To this end, both realistic Brownian like dynamics and unrealistic Monte-Carlo dynamics [41] that allow for the speeding up of the sampling of conformation space are used to generate conformations. Provided a distance below which two loci are considered to be in contact, contact pairs are then identified for every conformation, and the resulting statistics provides the contact map (Fig. 2).

2.2.2. Reproducing TAD formation and the decay of contact frequencies as a function of the genomic distance separating chromosomal loci

In the spirit of the RB approach proposed by Alber and co-workers [20,40], TB approaches consist of generating a multitude of conformations. These conformations can then be used to compute statistical properties, such as the mean and standard deviation of distances between certain loci, and can then be compared with the experimental measurements (using for instance 3D-FISH). In contrast to the RB approaches, though, the hypothesis and parameters that are used to conceive an appropriate polymer model are as instructive as the conformations that are generated.

The first genome-wide contact map of the human genome was "demuxed" using a simple model of homogeneous polymers in which all loci behaved in the same way [35]. The observed law of contact frequencies as a function of the distance s that separates two loci, $P(s) \sim s^{-\alpha}$ with $\alpha \approx 1$, was then explained as the result of a fractal organization of the polymer chains – such fractal organization corresponds to the long-living metastable states that emerge during the condensation of homogeneous polymers into globules [42] due, for instance, to confinement. More recent analyses of contact maps, together with more complex polymer modeling, instead suggest that specific interactions between loci are needed to both reproduce TADs [43,44] and explain the various laws that can be found for $P(s)$ with, in particular, an exponent α that varies between chromosomes [43]. In addition, attractive interactions between loci seem to be necessary to quantitatively reproduce the statistics of distance properties between loci, which can be measured using *in situ* fluorescent visualization techniques [45,43,24,46].

2.2.3. Devising predictive models

In addition to the possibility of designing specific models that are aimed at matching specific situations [22–24], TB approaches offer the possibility of devising predictive computational models [47,24]. In this regard, the recent work of Giorgetti and co-workers [24] provides an inspiring example of the potentiality of these frameworks. This work reports a bead-on-a-string polymer modeling of the human chromosome that is used to investigate the functional impact of TAD formation within the X-inactivation center. In the model, beads are able to interact according to some potential whenever their relative distance is below a certain threshold. Most importantly, in contrast to most modeling works of this type, *no parameter was pre-defined*, except the coarse-graining scale (3 kb). Instead, all parameters were adjusted using an optimization procedure coming from the field of protein folding [48], which involved progressively adapting the parameters and periodically simulating the thermodynamic properties of the corresponding model such that the model converges to the desired contact map. Using this approach, several remarkable results have been found. For instance, a 32-nm diameter emerges as the optimal size of the beads, which supports the relevance of an *effective* ca. 30 nm "chromatin fiber" for the X-inactivation center during interphase. Next, a typical 80 nm range was found for the contact interaction between loci, which supports the idea that interactions between specific loci are mediated by large protein complexes (with CTCF and cohesin among the main determinants). Remarkably, the polymer model was predictive in the sense that the deletion of a crucial site, such as a CTCF binding site, had similar effects both in vivo and *in silico*.

2.2.4. The meaning of optimal model parameters

TB approaches are not only useful for providing insights into the three-dimensional organization of chromosomes, but they also raise important questions regarding the physical meaning of the parameters that are used for a proper demuxing of contact maps.

Referring back to the study by Giorgetti and co-workers [24], both the optimal diameter of the effective fiber and the range of interactions are consistent with current knowledge of chromatin and chromosomes. Next, both attractive and repulsing interactions appear to be required to recapitulate experimental contact maps. Large attracting energies are compatible with a bridging mechanism that is induced by large protein complexes [49], which is consistent with the CTCF enrichment of the strongest interacting sites [24]; non-specific (weak) interacting energies may instead indicate a global interaction network between domains with similar epigenomic profiles [44]. The rationale for repulsive interactions is less clear. They might, for instance, indicate a volume exclusion effect that results from the presence of large complexes that are involved in other specific interactions. If this were the case, it would suggest that TB approaches might provide a *bona fide* framework for capturing not only the properties of chromosome folding but also the presence of associated macromolecular complexes.

3. A matrix-based demultiplexing approach

As shown in Figs. 1 and 2, experimental contact maps are often made of overlapping interaction blocks that are larger than TADs, which indicates that although TADs may constitute a relevant decomposition of contact maps, other decompositions can be contemplated *a priori*. In particular, contact maps such as that shown in Fig. 3 call for the presence of overlapping interaction domains that are larger than TADs and occur at specific frequencies. In this context, our goal here is to present a matrix analysis that can be used to estimate these frequencies.

3.1. A decomposition into statistical association domains (SIDs)

We call "statistical interaction domains" (SIDs) all of the domains that can be defined on the same basis as TADs but where the borders, though they are identical to those used to define TADs, do not have to be consecutive along the chromosome (Fig. 3). SIDs are thus larger than TADs; they can actually be smaller if we further consider sub-TAD borders instead of TAD borders [50,51]. Most importantly, SIDs can overlap with a large SID that can contain several smaller SIDs (Fig. 3 right panel).

Using SIDs as a basis for demuxing contact maps, we aim to estimate their frequency of occurrence in a cell population by exploiting a linear relationship that exists, under certain approximations, between these frequencies and the counts in every tile of the contact map (Fig. 4). To this end, let f_I be the occurrence frequency of SID I and let $\vec{f} = (f_1, \ldots, f_M)$ be the corresponding vector of all frequencies – we consider an ensemble of M SIDs formed from a set of m borders $\vec{b} = (b_1, \ldots, b_m)$, so that $M = m(m-1)/2$. A SID being defined by a pair of borders, we further sort these SIDs and their frequency according to the following order of pairs of borders: $(b_1, b_2), \ldots, (b_1, b_m), (b_2, b_3), \ldots, (b_2, b_m), \ldots (b_{m-1}, b_m)$.

Next, let $\vec{t} = (t_1, \ldots, t_M)$ be a vector such that t_J is equal to the sum of contact frequencies in the tile J of the contact map (see left panel in Fig. 4A). A tile is defined here by a pair (b_i, b_j) of borders such that (b_i, b_j) indicates the matrix position of the upper left tile corner and (b_{i+1}, b_{j+1}) the position of the lower right corner; for instance, in Fig. 4A, the tiles 1, 2 and 3 are defined by the pairs $(b_1, b_1), (b_2, b_1)$ and (b_2, b_2). In this context, one can easily verify that there are as many tiles as SIDs, namely, $M = m(m-1)/2$.

By sorting the tiles according to the following order of border pairs, $(b_1, b_1), \ldots, (b_{m-1}, b_1), (b_2, b_2), \ldots (b_{m-1}, b_2), \ldots (b_{m-1}, b_{m-1})$, and supposing that overlapping SIDs are exclusive, which indicates that a given cell cannot simultaneously display a "SID within a SID", one can write the following multiplexing relationship that links tiles and SIDs together:

$$\vec{t} = M \cdot \vec{f} \tag{1}$$

Here, the entries m_{JI} of the matrix M indicate the contribution of SID I to tile J, such that we have $t_J = \sum_{I=1}^{I=M} m_{JI} f_I$. More precisely, m_{JI} corresponds to the number of contacts in tile J that are provided by SID I each time this SID is formed (Fig. 4A). In particular, $m_{JI} = 0$ if SID I does not overlap with tile J, as is the case for SID 3 and tile 1 in Fig. 4A.

The possibility of writing this multiplexing relationship relies on the tacit assumption that interactions within a SID are diluted throughout the SID and that they occur according to the same contact law $p(s)$ (see below). In this case, the contribution m_{JI} of SID I to tile J is given by

$$m_{JI} = \frac{1}{L_I} \times \sum_{(a,b)\in J} p_I(s_{ab}), \tag{2}$$

where for every pair (a, b) of loci in tile J, s_{ab} indicates their genomic distance along DNA and $p_I(s)$ is the expected contact frequency between the loci in SID I as a function of their genomic distance. The larger the SID is, the smaller m_{JI}, whereas the larger the tile is, the larger m_{JI}, with a greatest value of $m_{JI} = 1$ when the SID and the tile coincide.

Note that for a given SID I, the contact law $p_I(s)$ reflects the nature of the interactions between the loci within the SID. We further suppose, here, that the functional form of $p_I(s)$ is the same for all SIDs I. It is then important to realize that this contact law is

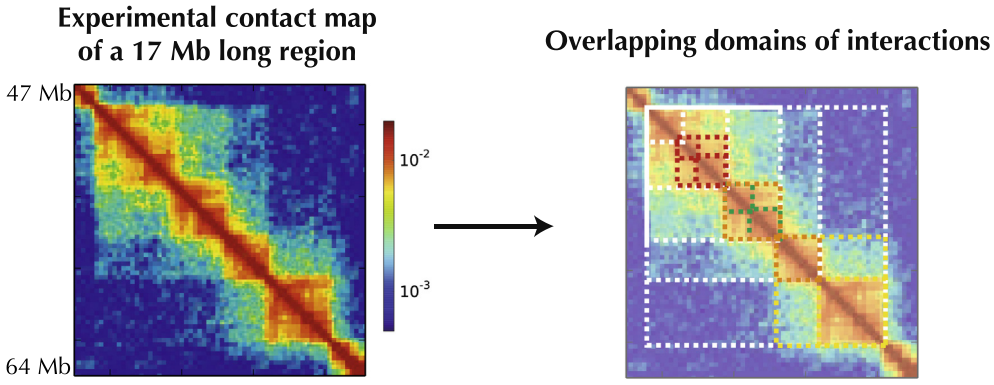

Fig. 3. Evidence of the presence of overlapping interaction domains in HiC contact maps. The input contact map (left panel) corresponds to a 17-Mb-long region of chromosome 4 in T47D cells. It was obtained using NcoI as the restriction enzyme, with a 20 kb resolution [25]. The initial raw map was normalized using ICE [26] and smoothed using a Gaussian filter with a 12 kb smoothing window. The resulting contact map clearly shows the presence of several overlapping domains of interactions, which are indicated on the right panel using dashed boxes.

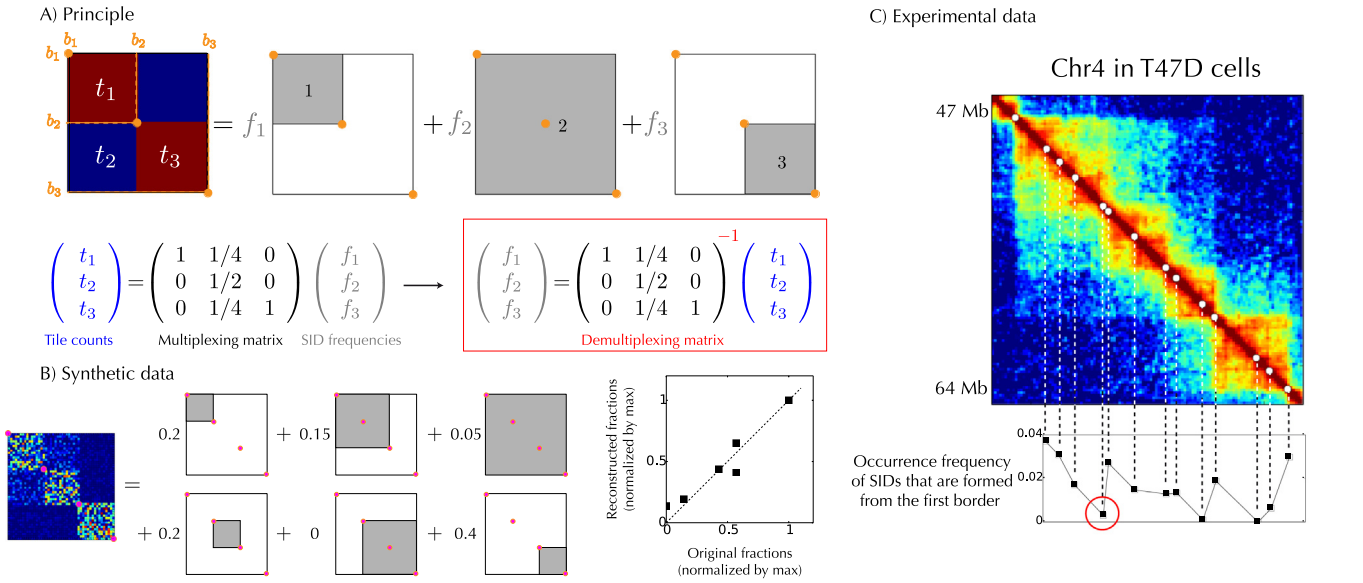

Fig. 4. A matrix-based demultiplexing approach. (A) Schematic representation of a contact map decomposition into the superimposition of statistical interaction domains (SIDs). Three SIDs, here on the right, occur with frequencies f_1, f_2 and f_3. They are formed from the combination of all pairs of borders (b_1, b_1), (b_2, b_1) and (b_2, b_2) that can be identified in the input contact map (orange points). SID combination results in a mosaic contact map made of tiles (corresponding here to t_1, t_2 and t_3), as it is often observed experimentally (see, e.g., Fig. 3). In this context and as indicated by the matrix relationship on the left, every sum t_1, t_2 and t_3 of the contact frequencies in the tiles can be written as a linear combination of the contact frequencies obtained from the SIDs (Eq. (3) in the text) – to this end, a normalized contact map for each SID is considered. In this linear relationship, the contribution of SID I to tile J is provided by Eq. (2), which in the case of a uniform contact probability is equal to the ratio between the tile area and the SID area (it is equal to 0 if these do not coincide). Inverting the relationship allows the occurrence frequency of SIDs to be determined from the tile counts (demuxing relationship on the right). (B) Application to a stylized dataset. The dataset was produced by the superimposition of 6 SIDs corresponding to 4 borders (pink points) and occurring with different frequencies. We additionally consider a noisy process to test the robustness of our method (see text for details). Right panel: reconstructed occurrence frequencies obtained from the demuxing relationship are in good agreement with input frequencies – frequencies were normalized by the maximum frequency. (C) Application to experimental data. By applying this demuxing approach to the data in Fig. 3, we obtain the occurrence frequencies of 91 SIDs corresponding to 14 borders (white points on contact map). Here, we indicate at the bottom of the map the relative occurrence frequencies of the 13 SIDs that can be formed using the first border. Large SIDs of more than 10 Mb can be as frequent as small ones, and in the case of two close borders, our method is capable of discriminating the presence of a "false positive" (red circle).

different from the contact frequency $P(s)$ that has been recently discussed in the literature (see above). Indeed, $P(s)$ corresponds to a contact frequency that is *averaged over all genomic loci and over a cell population* [35]. In contrast, $p_I(s)$ corresponds to the contact law that is expected *within a given genomic domain* (*SID I here*), *once the contact map has been demuxed*. $p_I(s)$ thus cannot be deduced using an averaging procedure over the contact map; it is an unknown function that must be determined self-consistently (see Section 4 for further details).

By inverting Eq. (1), we obtain a demuxing equation that allows computing SID frequencies ($\vec{f}$) from tile counts ($\vec{t}$). Namely, denoting the demuxing matrix $D = M^{-1}$, we have

$$\vec{f} = D \cdot \vec{t} (= M^{-1} \cdot \vec{t}) \tag{3}$$

This demuxing equation shows that under the assumptions that overlapping SIDs are exclusive, there is a unique set $\vec{f}$ of frequencies that correspond to (i) the tiles of the matrix and (ii) the contact frequency law $p_I(s)$. In other words, by performing a simple count analysis of the contact frequency in the tiles and considering simple assumptions on the form of $p_I(s)$ (see below), one can easily obtain SID frequencies.

3.2. Validation of the method: synthetic data

To test the validity of our approach, we apply it to the analysis of a well-controlled synthetic dataset (Fig. 4B). To this end, we consider a stylized contact map that is formed by the occurrence of 6 different SIDs, which correspond to 4 borders (Fig. 4B), with frequencies $\vec{f} = (0.2, 0.15, 0.05, 0.2, 0, 0.4)$. We also consider that interactions between loci within a given SID are all equivalent such that for a SID I with frequency f_I, an interaction between two loci occurs with the frequency f_I/L_I.

To address the robustness of our method, we further consider that the values of contact frequencies are drawn according to a random process. This random process aims to mimic the presence, in real experiments, of experimental artifacts or structural heterogeneities within SIDs. Specifically, for every contact in SID I, we consider that the original contributing frequency f_I/L_I is strongly distorted by an additional noise such that the measured value is equal to $\max\{0, (1+\eta)f_I/L_I\}$, where η is a random variable that takes its values with uniform distribution in $[-10, 10]$.

For this example, as a solution, we naturally consider an equal distribution of interactions within a SID such that we have $p_I(s) = 1/L_I$. In this context, denoting S_J as the sum of the contact frequencies in the tile J, we obtain $m_{JI} = S_J/L_I^2$ (the contribution of SID I to tile J is here simply equal to the ratio of their surfaces). As shown in the rightmost panel of Fig. 4B, Eq. (3) then leads to a good reconstruction of SID frequencies.

3.3. Experimental data: from TADs to SIDs

We now apply our methodology to the analysis of contact maps in T47D cells, for which HiC data were produced using HindIII and NcoI restriction enzymes at a resolution of 20 kb [25] and were normalized using the ICE procedure [26] by removing in particular the contribution from the diagonal [26]. We analyze more particularly a region of chromosome 4 that displays a well-defined superimposition of SIDs (Fig. 4C).

From this contact map, we first identify possible borders. To this end, we use a directional index [9] and identify the location where this index strongly varies from a positive value to a negative value.

We eventually obtain 14 plausible borders (white points on the contact map of Fig. 4C). From these borders, 91 SIDs (and tiles) can be defined, for which we can compute their frequency using Eqs. (2) and (3). To this end, just as in the case of synthetic data, we assume for simplicity that pair interactions within a given SID are equally distributed, thus implying $m_{jl} = S_j/L_l^2$. This approximation is equivalent to assume that a SID is an equilibrated polymer globule, with a polymer that has a small persistence length – note that this is a plausible scenario given that the persistence length of eukaryotic chromosomes may actually be very small, that is <30 nm [52,53] and that SID-like patterns are consistent with an equilibrated polymer where loci interact according to their epigenomic profile [44]; note also that the SIDs of Fig. 4C suggest a uniform distribution of interactions inside them rather than frequencies that decay with genomic distances (see also the high resolution maps of [3]). We finally stipulate that the obtained frequencies should all be positive – negative values occur here because of the sensitivity of small frequencies to noise. To this end, for each solution with a negative value, we set the frequency to zero and re-compute the solution using this constraint.

In this context, our analysis reveals that the formation of large SIDs can be as frequent as the formation of small SIDs. As an example, we indicate in Fig. 4C the frequency of occurrence of all of the SIDs that can be formed with the first border; 10-Mb-long SIDs are as frequent as 1-Mb-long SIDs. In addition, we observe that the frequency of SIDs is not a smooth function of their size, which corroborates that specificity plays a crucial role in the structuring of eukaryotic chromosomes.

4. Discussion and experimental considerations

To date, two generic structural methods have been developed to identify and quantify the multiple structures that are responsible for the contact maps obtained in 3C-based experiments. On the one hand, restraint-based (RB) approaches originate from the field of structural biology and rely on the satisfaction of distance restraints. On the other hand, thermodynamics-based (TB) approaches originate from the field of polymer physics and rely on a physical modeling of chromosome structuring.

Both approaches have advantages and drawbacks. RT methods can quickly provide information regarding the typical structuring of chromosomes. In addition, they do not assume any type of *a priori* regarding the underlying polymer models. They are thus well poised to handle the heterogeneity properties of chromatin in vivo. The possibility of properly quantifying the structural variability of chromosomes nevertheless remains an important challenge [21].

TB approaches provide a *bona fide*, likely realistic, description of chromosome structuring, which naturally includes the notion of structural variability. Importantly, they can be used to predict the impact of chromosome modification on the relative positioning of genes and their *cis*-regulatory sequences [47,24]. Their application for the multi-scale organization of a full eukaryotic chromosome remains an important challenge because of the time-consuming simulations and the intrinsic difficulty of predefining parameters for the polymer models.

4.1. A promising matrix-based framework with reasonable technical challenges

In this "structural" context, we have proposed a demuxing methodology based on a matrix analysis of contact maps that does not use explicit 3D structural information. Our goal was to provide a framework that is as simple as possible and allows for a quantitative analysis of heterogeneity properties of contact domains within chromosomes. This framework relies on the observation that contact maps resemble mosaics more than block diagonal matrices, with multiple domains that superimpose over a wide range of scales. In this context, we have extended the notion of topological associated domains (TADs) by considering the possibility of a superimposition of multiple interaction domains, called "statistical interaction domains" (SIDs), and we have presented a simple framework to compute their frequency in a cell population.

From a biological point of view, our preliminary results show that approximately 10-Mb-long domains can be as frequent as small ($\leqslant 1$ Mb) domains with frequencies that do not appear to be a simple function of the size of the SIDs. This raises the question of the physical origin of these SIDs. In the spirit of the scenario recently proposed by Jost et al. [44], SIDs might, for instance, correspond to different chromosomal conformations that result from a multi-stable behavior of chromosomes. Multi-stability would stem from the copolymer-like nature of chromosomes [44], which itself would result from the presence of a finite number of epigenomic domains [54,55,3] that are distributed throughout the genome and follow an intricate pattern. In the same spirit, SIDs might correspond to different outcomes (corresponding to different cells) of a polymer model that includes such self-attracting domains, but with the formation of loops that would stabilize a specific SID early on during the cell cycle. Such a looping-induced trapping effect could be set, for instance, during the mitotic phase, when the chromosomes are highly compacted. This would be consistent with a crucial structuring role played by cohesins and CTCF proteins [4,5]. This would also explain the apparent independence of the frequency of SIDs with respect to their size. In all cases, our framework should be useful for estimating the corresponding parameters of the TB models, both for the interaction energy between epigenomic profiles and for the site-specific energies that are related to the formation of loops, as these should be directly related to the SID frequencies.

From an operational point of view, having a normalized matrix in hand that provides an unbiased frequency of contacts between loci [31,26,56], let us recall here that our methodology relies on four pillars:

(a) The identification of borders using, e.g., directionality indexes [9].
(b) The assumption that contiguous domains can be defined on the same basis as TADs, or sub-TADs, but where the borders, though they are identical to those used to define TADs/sub-TADs, do not have to be consecutive along the chromosome.
(c) The assumption that overlapping SIDs are exclusive, which indicates that a given cell cannot simultaneously display a "SID within a SID".
(d) The choice of a generic law $p(s)$ for the decay, within SIDs, of the contact frequency as a function of the genomic distance separating loci.

The proper identification of borders (a) is not as crucial as it may be for defining TADs or sub-TADs. Indeed, as shown in Fig. 4C, the prediction of alternative "false positive" borders results in a frequency of occurrence that is negligible (red circle in bottom curve of Fig. 4C). In this regard, high-resolution maps [13,3] are expected to shed more light on the mechanism of SID formations.

Next, while the exclusive hypothesis (c) is a natural assumption that can be tested experimentally, the assumption that contiguous domains can be defined throughout the genome (b) requires further treatment. Indeed, in mammals, HiC interaction maps have revealed a ubiquitous two-compartment organization [35,3], which indicates that genomic domains belonging to different compartments poorly mix together. Importantly, this two-compartment pattern

is not compatible with the formation of SIDs along the genome; a SID that includes two domains that belong to different compartments would not verify the assumption that the interactions between loci of the SID are evenly distributed. A naïve application of our methodology to the whole genome thus leads to inconsistencies in the resulting frequencies, with the presence of large negative values. One possible way to solve this problem is to analyze compartments separately.

Finally, the choice of the decay law $p(s)$ (d) is a very interesting issue. It raises the question of the optimal law that should be considered to provide a reconstruction of the full matrix that is as similar to the contact maps as possible. Specifically, once the SID frequencies have been obtained, one can "reconstruct" the contact frequencies under the null hypothesis that SIDs occur with frequencies $\vec{f}$ and that the decay law within SIDs is given by $p(s)$. The values thus obtained can be compared to the original matrix, and a specific law $p(s)$ that maximizes the match can be searched for. This can be particularly useful to develop null models from which architectural elements can be determined [57].

4.2. Experimental considerations

Several major experimental challenges akin to our matrix-based demultiplexing of 3C related data need to be addressed. In particular, it is crucial to determine whether *different interaction domains* exist within a cell, or within a cell population, and to quantify their frequency of occurrence. It is also important to understand the functional impact of every interaction domain as well as the functional impact of having several of these interaction domains, knowing that many of them cannot occur at the same time.

A first possibility would be to confront our frequency predictions to single-cell HiC analysis [30]. The resolution achieved thus far with these techniques is not yet sufficient to obtain a quantitative assessment of domain frequencies. This stems from the low yield of recovery that these techniques can achieve and from the 3C principle itself, which allows only the detection of two events of ligation per hemi-site of restriction in a "frozen" conformation of the chromosome.

Another possibility would be to use fluorescent *in situ* hybridization (FISH) to analyze the relative position of different genomic sequences by microscopy. This technique has already been used to corroborate the presence of two compartments [35] and TADs [8], to characterize the statistical properties of distances between several loci [45,24,46] and to provide insights into the different organizations of a given domain within a cell/allele population [24]. In addition, recent advances in probe labeling and imaging have allowed high-resolution images to be obtained [58]. However, divergent conclusions are still obtained between FISH and 3C-related techniques, particularly when reaching sub-TAD resolution [59]. Moreover, whereas increasing the number of fluorescent labels could facilitate the interpretation, the possibilities of multi-channel imaging remain limited, particularly for high-resolution microscopy [60]. The development of novel powerful labeling methods [61] in combination with high-resolution microscopy [62,63] should hopefully help solve these issues.

While FISH methods allow for single cell analysis, they also rely on cell fixation. This implies that they cannot be used to answer questions regarding the stability of SIDs. In addition, although the visualization of specific loci at high resolution in time-lapse experiments remains challenging, the development of new labeling methods [64], particularly with the use of dedicated CRISPR-Cas9 systems [65], offers the possibility to follow the motion of genomic domains and should therefore permit the analysis of the stability of SIDs in living cells. CRISPR-Cas9 technology [66,67] also opens the way for a systematic manipulation of chromosome structural determinants. This should be crucial to pinpoint the mechanisms that underlie the specificity of border association.

Acknowledgments

We thank Marie Trussart for providing the chromosome structures of Figs. 1 and 2. I.J. is supported by an ATIP-Avenir grant (CNRS). The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007–2013)/ERC grant agreement n° 609989.

References

[1] Dekker, J., Rippe, K., Dekker, M. and Kleckner, N. (2002) Capturing chromosome conformation. Science 295 (5558), 1306–1311.
[2] De Wit, E. and De Laat, W. (2012) A decade of 3C technologies: insights into nuclear organization. Genes Dev. 26 (1), 11–24.
[3] Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., et al. (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159 (7), 1665–1680.
[4] Phillips, J.E. and Corces, V.G. (2009) CTCF: master weaver of the genome. Cell 137 (7), 1194–1211.
[5] Ong, C.-T. and Corces, V.G. (2014) CTCF: an architectural protein bridging genome topology and function. Nat. Rev. Genet..
[6] Phillips-Cremins, J.E. and Corces, V.G. (2013) Chromatin insulators: linking genome organization to cellular function. Mol. Cell 50 (4), 461–474.
[7] Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., et al. (2012) Three-dimensional folding and functional organization principles of the Drosophila genome. Cell 148 (3), 458–472.
[8] Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., et al. (2012) Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature 485 (7398), 381–385.
[9] Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., et al. (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485 (7), 376–380.
[10] Hou, C., Li, L., Qin, Z.S. and Corces, V.G. (2012) Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains. Mol. Cell 48 (3), 471–484.
[11] Naumova, N., Imakaev, M., Fudenberg, G., Zhan, Y., Lajoie, B.R., Mirny, L.A. and Dekker, J. (2013) Organization of the mitotic chromosome. Science 342 (6161), 948–953.
[12] Vietri Rudan, M., Barrington, C., Henderson, S., Ernst, C., Odom, D.T., Tanay, A. and Hadjur, S. (2015) Comparative HiC reveals that CTCF underlies evolution of chromosomal domain architecture. Cell Rep. 10 (8), 1297–1309.
[13] Gorkin, D.U., Leung, D. and Ren, B. (2014) The 3D genome in transcriptional regulation and pluripotency. Cell Stem Cell 14 (6), 762–775.
[14] Merkenschlager, M. and Odom, D.T. (2013) CTCF and cohesin: linking gene regulatory elements with their targets. Cell 152 (6), 1285–1297.
[15] Lanctôt, C., Cheutin, T., Cremer, M., Cavalli, G. and Cremer, T. (2007) Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. Nat. Rev. Genet. 8 (2), 104–115.
[16] Cavalli, G. and Misteli, T. (2013) Functional implications of genome topology. Nat. Struct. Mol. Biol. 20 (3), 290–299.
[17] Lucas, J.S., Zhang, Y., Dudko, O.K. and Murre, C. (2014) 3D trajectories adopted by coding and regulatory DNA elements: first-passage times for genomic interactions. Cell 158 (2), 339–352.
[18] Levine, M. (2014) The contraction of time and space in remote chromosomal interactions. Cell 158 (2), 243–244.
[19] Dekker, J. (2008) Gene regulation in the third dimension. Science 319 (5871), 1793–1794.
[20] Tjong, H., Gong, K., Chen, L. and Alber, F. (2012) Physical tethering and volume exclusion determine higher-order genome organization in budding yeast. Genome Res. 22 (7), 1295–1305.
[21] Trussart, M., Serra, F., Baù, D., Junier, I., Serrano, L. and Marti-Renom, M.A. (2015) Assessing the limits of restraint-based 3D modeling of genomes and genomic domains. Nucleic Acids Res.. in press.
[22] Junier, I., Dale, R.K., Hou, C., Képès, F. and Dean, A. (2012) CTCF-mediated transcriptional regulation through cell type-specific chromosome organization in the β-globin locus. Nucleic Acids Res. 40 (16), 7718–7727.
[23] Meluzzi, D. and Arya, G. (2013) Recovering ensembles of chromatin conformations from contact probabilities. Nucleic Acids Res. 41 (1), 63–75.
[24] Giorgetti, L., Galupa, R., Nora, E.P., Piolot, T., Lam, F., Dekker, J., et al. (2014) Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. Cell 157 (4), 950–963.
[25] Le Dily, F., Baù, D., Pohl, A., Vicent, G.P., Serra, F., Soronellas, D., et al. (2014) Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation. Genes Dev. 28 (19), 2151–2162.

[26] Imakaev, M., Fudenberg, G., McCord, R.P., Naumova, N., Goloborodko, A., Lajoie, B.R., et al. (2012) Iterative correction of HiC data reveals hallmarks of chromosome organization. Nat. Methods 9 (10), 999–1003.
[27] De Laat, W. and Duboule, D. (2013) Topology of mammalian developmental enhancers and their regulatory landscapes. Nature 502 (7), 499–506.
[28] Andrey, G., Montavon, T., Mascrez, B., Gonzalez, F., Noordermeer, D., Leleu, M., et al. (2013) A switch between topological domains underlies HoxD genes collinearity in mouse limbs. Science 340 (6137), 1234167.
[29] Noordermeer, D., Leleu, M., Schorderet, P., Joye, E., Chabaud, F. and Duboule, D. (2014) Temporal dynamics and developmental memory of 3D chromatin architecture at Hox gene loci. eLife 3, e02557.
[30] Nagano, T., Lubling, Y., Stevens, T.J., Schoenfelder, S., Yaffe, E., Dean, W., et al. (2013) Single-cell HiC reveals cell-to-cell variability in chromosome structure. Nature 502 (7), 59–64.
[31] Yaffe, E. and Tanay, A. (2011) Probabilistic modeling of HiC contact maps eliminates systematic biases to characterize global chromosomal architecture. Nat. Genet. 43 (11), 1059–1065.
[32] Hu, M., Deng, K., Qin, Z. and Liu, J.S. (2013) Understanding spatial organizations of chromosomes via statistical analysis of HiC data. Quant. Biol..
[33] Rosa, A. and Zimmer, C. (2014) Computational models of large-scale genome architecture. International review of cell and molecular biology 307, 275–349.
[34] Marti-Renom, M.A. and Mirny, L.A. (2011) Bridging the resolution gap in structural modeling of 3D genome organization. PLoS Comput. Biol. 7 (7), e1002125.
[35] Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., et al. (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326 (5950), 289–293.
[36] Baù, D., Sanyal, A., Lajoie, B.R., Capriotti, E., Byron, M., Lawrence, J.B., et al. (2011) The three-dimensional folding of the α-globin gene domain reveals formation of chromatin globules. Nat. Struct. Mol. Biol. 18 (1), 107–114.
[37] Rousseau, M., Fraser, J., Ferraiuolo, M.A., Dostie, J. and Blanchette, M. (2011) Three-dimensional modeling of chromatin structure from interaction frequency data using Markov chain Monte Carlo sampling. BMC Bioinformatics 12 (1), 414.
[38] Ay, F., Bunnik, E.M., Varoquaux, N., Bol, S.M., Prudhomme, J., Vert, J.-P., et al. (2014) Three-dimensional modeling of the *P. falciparum* genome during the erythrocytic cycle reveals a strong connection between genome architecture and gene expression. Genome Res. 24 (6), 974–988.
[39] Hu, M., Deng, K., Qin, Z., Dixon, J., Selvaraj, S., Fang, J., et al. (2013) PLoS Comput. Biol. 9 (1), e1002893.
[40] Kalhor, R., Tjong, H., Jayathilaka, N., Alber, F. and Chen, L. (2012) Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. Nat. Biotechnol. 30 (1), 90–98.
[41] Binder, K. and Heermann, D.W. (2010) Monte Carlo Simulation in Statistical Physics (Vol. 0). Monte Carlo Simulation in Statistical Physics: An Introduction.
[42] Grosberg, A., Rabin, Y., Havlin, S. and Neer, A. (1993) Crumpled globule model of the three-dimensional structure of DNA. EPL 23 (5), 373–378.
[43] Barbieri, M., Chotalia, M., Fraser, J., Lavitas, L.-M., Dostie, J., Pombo, A. and Nicodemi, M. (2012) Complexity of chromatin folding is captured by the strings and binders switch model. Proc. Natl. Acad. Sci. 109 (40), 16173–16178.
[44] Jost, D., Carrivain, P., Cavalli, G. and Vaillant, C. (2014) Modelling epigenome folding: formation and dynamics of topologically associated chromatin domains. Nucleic Acids Res. 42 (15), 9553–9561.
[45] Mateos-Langerak, J., Bohn, M., de Leeuw, W., Giromus, O., Manders, E.M.M., Verschure, P.J., et al. (2009) Spatially confined folding of chromatin in the interphase nucleus. Proc. Natl. Acad. Sci. 106 (10), 3812–3817.
[46] Tark-Dame, M., Jerabek, H., Manders, E.M.M., Heermann, D.W. and van Driel, R. (2014) Depletion of the chromatin looping proteins CTCF and cohesin causes chromatin compaction: insight into chromatin folding by polymer modeling. PLoS Comput. Biol. 10 (10), e1003877.
[47] Wong, H., Marie-Nelly, H., Herbert, S., Carrivain, P., Blanc, H., Koszul, R., et al. (2012) A predictive computational model of the dynamic 3D interphase yeast nucleus. Curr. Biol. 22 (20), 1881–1890.
[48] Norgaard, A.B., Ferkinghoff-Borg, J. and Lindorff-Larsen, K. (2008) Experimental parameterization of an energy function for the simulation of unfolded proteins. Biophys. J. 94 (1), 182–192.
[49] Junier, I., Martin, O. and Képès, F. (2010) Spatial and topological organization of DNA chains induced by gene co-localization. PLoS Comput. Biol. 6 (2), e1000678.
[50] Berlivet, S., Paquette, D., Dumouchel, A., Langlais, D., Dostie, J. and Kmita, M. (2013) Clustering of tissue-specific sub-TADs accompanies the regulation of HoxA genes in developing limbs. PLoS Genet. 9 (12), e1004018.
[51] Phillips-Cremins, J.E., Sauria, M.E.G., Sanyal, A., Gerasimova, T.I., Lajoie, B.R., Bell, J.S.K., et al. (2013) Architectural protein subclasses shape 3D organization of genomes during lineage commitment. Cell 153 (6), 1281–1295.
[52] Ben-Haïm, E., Lesne, A. and Victor, J.-M. (2001) Chromatin: a tunable spring at work inside chromosomes. Phys. Rev. E 64 (5), 51921.
[53] Hajjoul, H., Mathon, J., Ranchon, H., Goiffon, I., Mozziconacci, J., Albert, B., et al. (2013) High-throughput chromatin motion tracking in living yeast reveals the flexibility of the fiber throughout the genome. Genome Res. 23 (11), 1829–1838.
[54] Filion, G.J., van Bemmel, J.G., Braunschweig, U., Talhout, W., Kind, J., Ward, L.D., et al. (2010) Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. Cell 143 (2), 212–224.
[55] Roudier, F., Ahmed, I., Bérard, C., Sarazin, A., Mary-Huard, T., Cortijo, S., et al. (2011) Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. EMBO J. 30 (10), 1928–1938.
[56] Cournac, A., Marie-Nelly, H., Marbouty, M., Koszul, R. and Mozziconacci, J. (2012) Normalization of a chromosomal contact map. BMC Genomics 13, 436.
[57] Ay, F., Bailey, T.L. and Noble, W.S. (2014) Statistical confidence estimation for HiC data reveals regulatory chromatin contacts. Genome Res. 24 (6), 999–1011.
[58] Giorgetti, L., Piolot, T. and Heard, E. (2015) High-resolution 3D DNA FISH using plasmid probes and computational correction of optical aberrations to study chromatin structure at the sub-megabase scale. Methods Mol. Biol. 1262 (Chapter 3), 37–53.
[59] Williamson, I., Berlivet, S., Eskeland, R., Boyle, S., Illingworth, R.S., Paquette, D., et al. (2014) Spatial genome organization: contrasting views from chromosome conformation capture and fluorescence in situ hybridization. Genes Dev. 28 (24), 2778–2791.
[60] Bates, M., Huang, B., Dempsey, G.T. and Zhuang, X. (2007) Multicolor super-resolution imaging with photo-switchable fluorescent probes. Science 317 (5845), 1749–1753.
[61] Ma, H., Naseri, A., Reyes-Gutierrez, P., Wolfe, S.A., Zhang, S. and Pederson, T. (2015) Multicolor CRISPR labeling of chromosomal loci in human cells. Proc. Natl. Acad. Sci. 112 (10), 3002–3007.
[62] Huang, B., Babcock, H. and Zhuang, X. (2010) Breaking the diffraction barrier: super-resolution imaging of cells. Cell 143 (7), 1047–1058.
[63] Markaki, Y., Smeets, D., Fiedler, S., Schmid, V.J., Schermelleh, L., Cremer, T. and Cremer, M. (2012) The potential of 3D-FISH and super-resolution structured illumination microscopy for studies of 3D nuclear architecture: 3D structured illumination microscopy of defined chromosomal structures visualized by 3D (immuno)-FISH opens new perspectives for studies of nuclear architecture. BioEssays 34 (5), 412–426.
[64] Saad, H., Gallardo, F., Dalvai, M., Tanguy-le-Gac, N., Lane, D. and Bystricky, K. (2014) DNA dynamics during early double-strand break processing revealed by non-intrusive imaging of living cells. PLoS Genet. 10 (3), e1004187.
[65] Chen, B., Gilbert, L.A., Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.-W., et al. (2013) Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. Cell 155 (7), 1479–1491.
[66] Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A. and Zhang, F. (2013) Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. 8 (11), 2281–2308.
[67] Doudna, J.A. and Charpentier, E. (2014) The new frontier of genome engineering with CRISPR-Cas9. Science 346 (6213), 1258096.